



Texas Pay-for-Quality (P4Q) Program Technical Specifications

Measurement Periods:

Implementation Year: Calendar Year 2014

Baseline Year: Calendar Year 2013

**The Institute for Child Health Policy
University of Florida**

**The External Quality Review Organization
for Texas Medicaid Managed Care and CHIP**

**Revised:
August 11, 2014**



Table of Contents

| | |
|---|----|
| Overview | 1 |
| Concepts | 1 |
| Quality Measures | 3 |
| Data Periods..... | 4 |
| The Measures..... | 4 |
| Data Sources | 11 |
| Data Preparation..... | 11 |
| Risk Adjustment | 11 |
| Methodology Overview | 11 |
| Risk Adjustment Variables | 13 |
| Using Hybrid Data Instead of Administrative Data | 13 |
| Calculation of Gap Closure..... | 14 |
| Calculation of Points | 15 |
| Raw Points..... | 15 |
| Adjusted Points..... | 16 |
| Converting Points to Dollars | 18 |
| Issues in P4Q Design and Implementation | 19 |
| Risk Adjustment | 19 |
| Statistical Significance in Pay-for-Quality (P4Q) | 19 |
| Intra-Year Updates..... | 21 |
| Appendix 1. Gap Closure and the Partial Adjustment Model..... | 22 |
| Appendix 2. Dental P4Q Earn Back Model..... | 24 |

Overview

This document presents the technical specifications for the Texas Pay-for-Quality (P4Q) Program for the health plans in STAR, CHIP, and STAR+PLUS and the dental plans in Medicaid and CHIP. The report begins with an overview of the concepts that form the basis of the P4Q program followed by a listing of the quality measures used across all programs and plans. The remainder of the report then discusses:

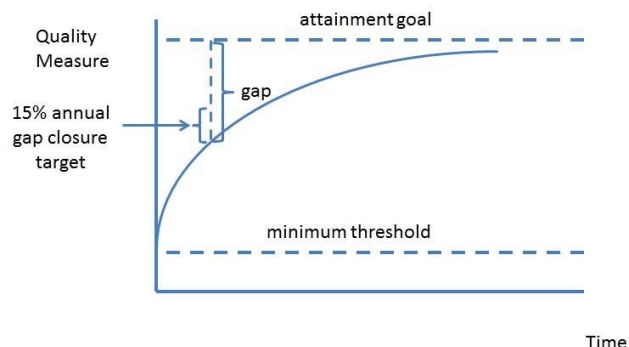
- Data sources
- Data preparation
- Risk adjustment
- Calculation of quality points, and
- Conversion of points into dollar amounts

The report concludes with a discussion of some important issues surrounding both the design and implementation of the P4Q Program. Two appendices provide additional detail on particular aspects of the P4Q program.

Concepts

The Texas P4Q Program is based on the concept of incremental improvement where each plan (1) is incentivized to improve its own quality performance each year and (2) is evaluated based on its success in achieving such improvement. By evaluating each plan based on its own performance, all plans have an opportunity to succeed in the program. Individual plans with lower quality scores are less likely to view targeted quality standards as beyond their reach and plans with higher quality scores remain involved in maintaining and improving their quality performance.

Figure 1
How Incremental Improvement Works



To implement the concept of incremental improvement, a gap closure approach to each P4Q quality measure is used, as depicted in Figure 1. In the gap closure approach, a *minimum threshold* is set which is the minimum quality score where plans become eligible to earn positive quality points. Plans with scores below the minimum threshold for a measure will not be eligible to earn

positive points for incremental improvement, but will not be penalized as long as they show year-to-year improvement in the measure. Also, an *attainment goal* is specified which represents a recognized level of excellence for the specific quality measure. This attainment goal is seen as the end product of a series of incremental improvements across time. Both minimum thresholds and attainment goals can be

different for different measures based on current plan performance, nationally recognized levels of excellence, factors specific to the Texas environment, and other factors.

Incremental improvement is defined in relation to the gap between the plan's current level of quality performance and the attainment goal for the measure (see the vertical distance labelled "gap" in Figure 1). Each plan is expected to close their gap by 15 percent each year. This 15 percent annual gap closure target illustrated in Figure 1 represents the amount of improvement in the quality measure that the plan should target for the current year to be on track for eventually achieving the attainment goal.

It is important to emphasize that this 15 percent annual gap closure target is not referring to a 15 percentage point annual increase in the measure. Rather, it refers to 15 percent of the arithmetic difference between the attainment goal and the plan's current level of performance. For example, if a health plan's current performance is 60 percent for the measure and the attainment goal is 70 percent, the gap is 70 percent - 60 percent or 10 percentage points. Correspondingly, 15 percent of the gap equals 0.15×10 percentage points or 1.5 percentage points. *The 15 percent annual gap closure target will be achieved if the plan improves from 60 percent to 61.5 percent on the measure.*

While higher scores on the HEDIS measures correspond to higher quality, *lower* PPE expenditures are indicative of higher quality. For this reason, improvement for the PPE measures is defined by a reduction in PPE expenditures. For example, if a plan currently incurs PPE expenditures of \$3,000 and its attainment goal is \$2,000, then the gap is -\$1,000 and 15 percent of this gap is -\$150. The 15 percent annual gap closure will be achieved if the plan reduces its PPE expenditure measure from \$3,000 to \$2,850. (Note: PPE expenditures in the P4Q program are calculated using standardized resource units in place of paid amounts for reasons discussed below. This example is presented in dollars for purposes of clarity.)

This gap closure approach recognizes that incremental improvement in quality will likely become more difficult as quality performance improves over time. When a plan first implements strategies designed to improve the quality of care, the improvement is likely to be substantial. Once this initial improvement is achieved, however, further improvement will require additional initiatives that will likely yield diminishing returns. The gap closure approach takes diminishing returns into account since 15 percent of the gap will decrease as the plan's quality performance increases and the overall gap decreases. This is shown in Figure 1 by the smooth curve that increases at a decreasing rate. This curve represents what would happen if the plan exactly achieved its 15 percent gap closure each year. Of course, no plan is likely to achieve exactly 15 percent gap closure each year, so the actual time path of quality for plans will vary by being higher in some years and lower in others.

The gap closure approach described here is patterned after an econometric model of partial adjustment to equilibrium¹ as described in Appendix 1.

1 Dougherty, Christopher (2012) EC220 - Introduction to econometrics (chapter 11). [Teaching Resource: London School of Economics, <http://learningresources.lse.ac.uk/137/>]

Quality Measures

This section presents the different quality measures that are used for the P4Q program across the STAR, CHIP, STAR+PLUS, and dental programs. These measures were chosen based on their importance for clinical care and population health, areas of emphasis as defined by the State of Texas and the Texas Legislature, and the likelihood that sufficient data would be available to produce stable estimates at the health plan level from year to year. All measures were chosen based on extensive discussions (and revisions) between HHSC, the plans, and the EQRO and analytic simulations using historical data to show how the measures would perform in the P4Q framework.

Some quality of care measures are Healthcare Effectiveness Data Information Set (HEDIS®) measures. For these measures, plans must have a minimum of 30 eligible enrollees for the measure in order to have the measure included in their P4Q calculations. “Eligible” is defined here as meeting the specified criteria for inclusion in the denominator of a HEDIS measure. If a plan has fewer than 30 enrollees for a given measure, that measure is not included in that plan’s P4Q calculations and a proportional adjustment is made to the plan’s quality points as described below to compensate for the missing measure.

Some HEDIS quality measures have several components (e.g., separate components for different age groups). The components that comprise a given measure are weighted such that each measure is given a total weight of +1.0. Typically, this results in measures comprised of two components having weights of +0.5 per component, measures comprised of three components having weights of +0.33 per component, and so forth.

Other STAR and STAR+PLUS quality measures are based on the 3M Health Information’s Potentially Preventable Events (PPEs), which include potentially preventable admissions (PPAs), readmissions (PPRs), and emergency department visits (PPVs). The precise 3M measures used in the P4Q program are PPA, PPR, and PPV expenditures per 1,000 member-months, with expenditures defined using standardized resource units rather than the actual amounts paid in order to remove market-based variations beyond the plans’ control. The quality measures used for the dental plans are based on state-defined measures and state-defined results.

Tables 1-7 present the Texas P4Q Program quality measures for the STAR, CHIP, STAR+PLUS, and dental programs, along with their minimum thresholds, attainment goals, and weights. Tables 1 and 2 present the health plan and dental plan P4Q measures, respectively, while Tables 3-7 present the measures with the sources of their associated minimum thresholds, attainment goals, and weights. For the CY2014 computations, ICHP will use HEDIS 2014 percentiles to set thresholds and attainment goals. For PPEs, the CY2013 program level risk adjusted standardized resource unit expenditures will be used.

The minimum thresholds for the STAR, CHIP, and STAR+PLUS programs are set using HEDIS 2014 national NCQA Medicaid 50th percentiles for HEDIS measures (with the exception of (1) the PPC prenatal/postpartum measure and (2) the HbA1c measure, which were set at the HEDIS 25th percentile) while attainment goals are set at the HEDIS 2014 NCQA 90th percentiles. Thresholds for PPEs are set as

the average program wide risk adjusted expenditures (based on standardized resource units) calculated using Texas statewide data while attainment goals are set at (1) a 25 percent reduction from each plan's starting value for those plans that start below the state mean or (2) a 25 percent reduction from the state mean for those plans that start at or above the state mean.

The quality measures for the dental P4Q program are presented in Tables 6 and 7. The minimum thresholds and the attainment goals for these dental measures were set based on discussions with HHSC and the dental plans. These measures are not risk adjusted.

Data Periods

The P4Q program uses data from the two most recent calendar years for purposes of calculating gap closures as described below.

The Measures

The quality measures for the 2014 STAR, STAR+PLUS, and CHIP P4Q programs are listed in Table 1. The quality measures for the dental P4Q program are listed in Table 2.

Table 1
2014 STAR, STAR+PLUS, and CHIP P4Q Program Measures

| Measure | Source | STAR | STAR+PLUS | CHIP |
|---|--------|------|-----------|------|
| Well-Child Visits 3, 4, 5, & 6 yr olds | HEDIS | X | | X |
| PPC – Postpartum care | HEDIS | X | | |
| PPC - Timeliness of prenatal care | HEDIS | X | | |
| Adolescent Well Care | HEDIS | X | | X |
| Antidepressant Med. Mgt. - Acute Phase | HEDIS | | X | |
| Antidepressant Med. Mgt. - Continuation Phase | HEDIS | | X | |
| HbA1c control <8 | HEDIS | | X | |
| Potentially Preventable Admissions | | | | |
| -- Risk Adjusted Expenditures /1,000 MM | 3M | X | X | X |
| Potentially Preventable Re-Admissions | | | | |
| -- Risk Adjusted Expenditures/ 1,000 MM | 3M | X | X | |
| Potentially Preventable ED Visits | | | | |
| -- Risk Adjusted Expenditures/ 1,000 MM | 3M | X | X | X |

Table 2
Texas Pay for Quality Measures – Medicaid and CHIP Dental Programs

| Measure | Medicaid | CHIP |
|---|----------|------|
| Percent of members (1 - 20 years old) enrolled for at least 11 of the past 12 months who had at least one preventive dental service during the measurement year | X | |
| Percent of new members (6 mo - 20 yrs) receiving at least one THSteps Dental Checkup within 90 days of enrollment | X | |
| THSteps Care Measures | X | |
| a. Percent of members (6 months - 20 years) receiving one THSteps Dental Checkup per year | | |
| b. Percent of members (6 months - 20 years) receiving two THSteps Dental Checkup per year | | |
| Sealant Measure | X | |
| a. percent of members (6-9) receiving at least one sealant | | |
| b. percent of members (10-14) receiving at least one sealant | | |
| Preventive Dental Services | | X |
| percent of members (1 - 18 yrs) with no more than one month gap enrollment who had at least one preventive dental service | | |
| Annual Dental Visit | | X |
| a. percent of members who had at least one annual dental visit by age group: | | |
| b. 2-3 yrs. | | |
| c. 4-6 yrs. | | |
| d. 7-10 yrs. | | |
| e. 11-14 yrs. | | |
| f. 15-18 yrs. | | |
| Sealant Measure | | X |
| percent of members receiving at least one sealant by age group: | | |
| a. 6-9 yrs. | | |
| b. 10-14 yrs. | | |

Tables 3-7 show the health and dental plan measures with their corresponding minimum thresholds, attainment goals, and weights.

Table 3
Texas Pay for Quality Measures – STAR Program

| Source | Measure | Minimum Threshold | Attainment Goal | Weight |
|---------------|---|---|--|---------------|
| HEDIS | Well-Child Visits 3, 4, 5, & 6 yr olds | HEDIS 2014 50 th percentile | HEDIS 2014 90 th percentile | 1.0 |
| HEDIS | PPC – Postpartum care | HEDIS 2014 25 th percentile | HEDIS 2014 90 th percentile | 0.5 |
| | PPC - Timeliness of prenatal care | | | 0.5 |
| HEDIS | Adolescent Well Care | HEDIS 2014 50 th percentile | HEDIS 2014 90 th percentile | 1.0 |
| 3M | Potentially Preventable Admissions -- Risk Adjusted Expenditures /1,000 MM | CY2013 Mean Program wide PPA Expenditures | 25 percent reduction | 1.0 |
| 3M | Potentially Preventable Re-Admissions -- Risk Adjusted Expenditures/ 1,000 MM | CY2013 Mean Program wide PPR Expenditures | 25 percent reduction | 1.0 |
| 3M | Potentially Preventable ED Visits -- Risk Adjusted Expenditures/ 1,000 MM | CY2013 Mean Program wide PPV Expenditures | 25 percent reduction | 1.0 |

Table 4
Texas Pay for Quality Measures – CHIP Program

| Source | Measure | Minimum Threshold | Attainment Goal | Weight |
|---------------|--|--|---|---------------|
| HEDIS | Well-Child Visits 3, 4, 5, & 6 yr olds | HEDIS 2014 50 th percentile | HEDIS 2014 90 th percentile | 1.0 |
| HEDIS | Adolescent Well Care | HEDIS 2014 50 th percentile | HEDIS 2014 90 th percentile | 1.0 |
| 3M | Potentially Preventable Admissions -- Risk Adjusted Expenditures/ 1,000 MM | Mean Program- wide PPA Expenditures | 25 percent reduction | 1.0 |
| 3M | Potentially Preventable ED Visits -- Risk Adjusted Expenditures/ 1,000 MM | Mean Program-wide PPV Expenditures | 25 percent reduction | 1.0 |

Table 5
Texas Pay for Quality Measures – STAR+PLUS Program

| Source | Measure | Minimum Threshold | Attainment Goal | Weight |
|---------------|--|---|---|---------------|
| HEDIS | Antidepressant Med. Mgt. | | | |
| | a) Acute Phase | HEDIS 2014 50 th percentile | HEDIS 2014 90 th percentile | 0.5 |
| | b) Continuation Phase | | | 0.5 |
| HEDIS | HbA1c control <8 | HEDIS 2014 25 th percentile | HEDIS 2014 90 th percentile | 1.0 |
| 3M | Potentially Preventable Admissions -- Risk Adj Expenditures/1000 MM | Mean Program-wide PPV Expenditures | 25 percent reduction | 1.0 |
| 3M | Potentially Preventable Re- Admissions -- Risk Adj Expenditure/1000 MM | Mean Program-wide PPV Expenditures | 25 percent reduction | 1.0 |
| 3M | Potentially Preventable ED Visits -- Risk Adj Expenditures/1000 MM | Mean Program-wide PPV Expenditures | 25 percent reduction | 1.0 |

Table 6
Texas Pay for Quality Measures – Medicaid Dental Program

| Measure | Minimum Threshold | Attainment Goal | Weight |
|--|------------------------------|----------------------------|---------------|
| Percent of members (1 - 20 years old) enrolled for at least 11 of the past 12 months who had at least one preventive dental service during the measurement year. | 64.8 | 85 | 1.0 |
| Percent of new members (6 mo - 20 yrs) receiving at least one THSteps Dental Checkup within 90 days of enrollment | 18 | 40 | 1.0 |
| THSteps Care Measures ² (Composite rate= (1.0*rate of 2 visits)+(0.5*rate of 1 visit), threshold and attainment goal are based on composite rate) | 46.8 | 65 | 1.0 |
| a. Percent of members (6 months - 20 years) receiving exactly one THSteps Dental Checkup per year | | | |
| b. Percent of members (6 months - 20 years) receiving exactly two THSteps Dental Checkup per year | | | |
| Sealant Measure (each sub-measure weighted equally) | | | |
| a. percent of members (6-9) receiving at least one sealant | 18.9 | 35 | 0.5 |
| b. percent of members (10-14) receiving at least one sealant | 11.7 | 30 | 0.5 |

² The two THSteps submeasures are combined into a single composite measure to facilitate interpretation since their construction allows a possible inverse interdependency between the two sub measures.

Table 7
Texas Pay for Quality Measures – CHIP Dental Program

| Measure | Minimum Threshold | Attainment Goal | Weight |
|---|------------------------------|----------------------------|---------------|
| Preventive Dental Services | | | |
| percent of members (1 - 18 yrs) with no more than one month gap enrollment who had at least one preventive dental service | 57.6 | 80 | 1.0 |
| Annual Dental Visit | | | |
| percent of members who had at least one annual dental visit by age group: | | | |
| a) 2-3 yrs. | 55.8 | 80 | 0.2 |
| b) 4-6 yrs. | 63 | 88 | 0.2 |
| c) 7-10 yrs. | 64.8 | 90 | 0.2 |
| d) 11-14 yrs. | 60.3 | 85 | 0.2 |
| e) 15-18 yrs. | 52.2 | 75 | 0.2 |
| Sealant Measure | | | |
| percent of members receiving at least one sealant by age group: | | | |
| a) 6-9 yrs. | 15.3 | 30 | 0.5 |
| b) 10-14 yrs. | 9 | 25 | 0.5 |

Data Sources

Data are obtained from STAR, CHIP, STAR+PLUS, and dental enrollment and encounter records. These records are used to produce (1) member level compliance or outcomes for each measure and (2) clinical and demographic information for each member.

- For all HEDIS and dental measures, data are member-based compliance (0/1) indicators for eligible members. Compliant members from the denominator (eligible members) are those that are counted in the numerator for rate calculation.
- For all 3M PPE measures, data are event-based observances of potentially preventable status of eligible admissions (PPA, PPR) or emergency department visits (PPV). Events from the qualified pool are flagged as potentially preventable, or not (0/1). Expenditures are calculated using standardized resource units rather than the actual paid amounts from the encounter records in order to remove the influence of extraneous market-based factors (e.g., geographic price level, provider market concentration, etc.) beyond the plans' control. These standardized resource units were calculated based on both Texas-specific and 3M national data using algorithms based on 3M's methods.

Data Preparation

The P4Q program relies on NCQA-approved HEDIS software and 3M PPE software for calculating eligibility and compliance. The EQRO ensures that data input to this software conforms to the specifications as outlined by the software vendors. Eligible populations of members or events are determined by individual measure specification, and are not altered for P4Q.

The following section describes the PPE risk adjustment process as specified by 3M. At present, the HEDIS measures used in P4Q are **not** risk adjusted, nor are any additional risk adjustors beyond those used by 3M applied to the PPE measures.

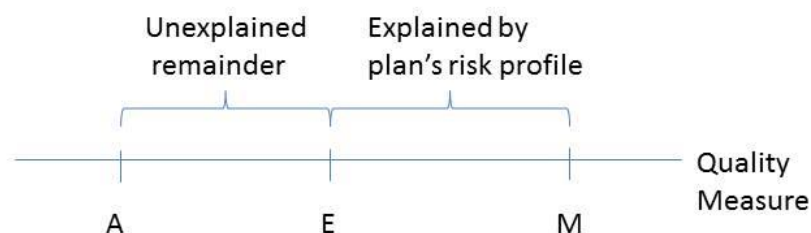
Risk Adjustment

Methodology Overview

Because of the lack of consensus on appropriate risk adjustors for the HEDIS measures combined with the on-going HHSC and MCO review and discussion of potential risk adjustment algorithms, the HEDIS measures in the P4Q program are not risk adjusted. However, the PPAs, PPVs, and PPRs are risk adjusted for health status using the procedures specified by 3M. The remainder of this section describes the rationale behind risk adjustment and how it works in relationship to the PPE measures.

The rationale behind risk adjustment is shown in Figure 2 where a plan's actual score, A, on a quality measure is shown along with the mean quality score across all plans in the program, M. The goal of risk adjustment is to use the statewide experience across all plans in the program to predict the expected

Figure 2 – How Risk Adjustment Works



Explained by plan's risk profile - portion of plan's performance difference due to factors beyond the plan's control.

Unexplained remainder - portion of plan's performance difference more likely related to factors under the plan's control (e.g., provider quality and availability, prevention policies, etc.)

quality score, E, that would occur given the plan's risk profile, i.e., those characteristics of the plan's enrollee population that influence outcomes and quality (e.g., diagnostic mix, severity, and mental health/substance abuse) that are predominately outside of the health plan's control. In Figure 2, the plan's expected quality score, E, divides the gap between plan's actual performance, A, and the mean quality score across all plans, M, into two parts. Moving right to left in Figure 2, the distance from M to E is the portion of the gap explained by the plan's risk profile. The second part (from E to A) is the unexplained remainder, which is presumed to include the effects of those factors that are under the plan's control, such as the availability and quality of the plan's provider network, the plan's quality assurance policies, and so forth. Risk adjustment seeks to remove the effects of the specified factors that influence the quality of care that are beyond the plan's control. The presumption is that the remaining difference between the health plan's actual and expected performance better reflects the health plan's success in improving the quality of care.

The risk-adjusted measure is calculated as the product of the actual-to-expected ratio (A/E) and the statewide mean, M, as shown in Figure 2.

Risk measure definitions can change from year-to-year as refinements are made to quality measurement systems and software. This can pose problems for incremental improvement calculations since such calculations rely on the year-to-year arithmetic changes in risk measures. These change calculations must be based on consistent definitions in both the baseline and performance years so that the calculations are meaningful. The EQRO will carefully monitor the changes in quality measure definitions in order to gauge measure comparability across years. When changes in quality measurement systems significantly reduce comparability across years, the EQRO will adjust measure definitions and/or gap closure calculations as necessary to ensure comparability. This may require the recalculation of baseline results using the new quality measure definitions. Should such adjustments or recalculations become necessary, the EQRO will communicate this information along with the new calculations to HHSC and the MCOs as soon as possible.

Risk Adjustment Variables

For the PPE expenditure measures, ICHP follows 3M's recommended categorical risk adjustment approach as implemented in the 3M PPE software. Instead of tallying events or paid amounts for PPAs, PPRs, and PPVs, however, ICHP uses standardized resource units for each plan in place of the amounts paid for each event. ICHP then uses the Texas statewide program data to create a norms file categorized by health status as described below. This norms file is the source of the expected PPE resource utilization to be used in the actual-to-expected ratio. As discussed above, PPE adjusted expenditures are calculated using standardized resource units rather than amounts paid in order to eliminate market- and provider-based variations in paid amounts. Refer to the 3M documentation for more details.

Using Hybrid Data Instead of Administrative Data

To date, all of the HEDIS measures used in the P4Q program have been based upon electronic administrative encounter data. To improve the accuracy of HEDIS measurement, the P4Q program will be moving to hybrid data based upon chart abstraction for most HEDIS measures. Therefore, in the future, the calculations described in this report will be based on these new hybrid data.

Calculation of Gap Closure

Each plan's performance on a given quality measure is based on the plan's year-to-year incremental improvement and how that incremental improvement compares to the target 15 percent gap closure as described above. To measure this incremental improvement, ICHP calculates each plan's actual gap closure for each quality measure in the P4Q program:

$$\text{actual gap closure} = \frac{\text{quality score}_t - \text{quality score}_{t-1}}{\text{attainment goal} - \text{quality score}_{t-1}}$$

The numerator of this actual gap closure calculation is the improvement in the quality score from the previous year to the current year. The denominator is the full amount of the gap, i.e., the difference between attainment goal and the plan's quality measure in the previous year.

An example of the calculation of the actual gap closure is presented in Figure 3. The score for Measure A improves from 57 percent in the year t-1 to 65 percent in year t for an improvement of 8 percent. The gap between the attainment goal for the measure and where the plan starts is 85 percent minus 57 percent, or 28 percent.

Figure 3
Calculating Actual Gap Closure

| Measure | Year t | Year t-1 | Target Closure | Attainment Goal | Actual Closure |
|-----------|--------|----------|----------------|-----------------|----------------|
| Measure A | 65% | 57% | 0.15 | 0.85 | +0.293 |

$$\text{Actual Closure} = \frac{\text{Year}_t - \text{Year}_{t-1}}{\text{Attainment Goal} - \text{Year}_{t-1}} = \frac{0.65 - 0.57}{0.85 - 0.57} = \frac{+0.08}{+0.28} = +0.293$$

In the example calculation shown in Figure 3, actual closure is +0.293, which exceeds the target gap closure of +0.15.

In general, the denominator of the actual closure calculation will be positive. However, once a plan exceeds the attainment goal, the actual closure must be multiplied by -1 so that the ratio has the appropriate sign.

Calculation of Points

Raw Points

For each measure, raw quality points are assigned based on the actual closure rate and how it compares to the target closure rate of 15 percent. Positive points are assigned for gap closure (increasing quality) while negative points are assigned for gap widening (decreasing quality). A maximum of +5 points are assigned and a minimum of -5 points are assigned for each measure.

To be eligible to earn positive points, a plan must be at or above the minimum threshold value for the given measure.

The relationship between actual gap closure and raw points assignment is shown in Table 8. When a plan meets or exceeds the attainment goal for a given measure, the plan receives the full +5 raw points. If the plan does not meet or exceed the attainment goal but scores above the minimum threshold and achieves actual gap closure of 15 percent or higher, the plan receives +4 points. If the plan does not meet or exceed the attainment goal, scores above the minimum threshold, and achieves actual gap closure between 0 and 15 percent, positive points are given on a “partial credit” basis in accordance with the ranges listed in Table 8. Zero points are assigned for gap closures below $\frac{1}{4}$ of the target gap closure, +1 points for $\frac{1}{4}$ - $\frac{1}{2}$ of the target gap closure, +2 points for $\frac{1}{2}$ - $\frac{3}{4}$ of the target gap closure, and +3 points for $\frac{3}{4}$ to 1 of the target gap closure below the attainment goal.

Table 8

Raw Points Assignment Based on Actual Gap Closure

| Positive Points | Gap Closure | Negative Points | Gap Widening |
|-----------------|--|-----------------|---------------------------|
| 5 | At or above attainment goal | -5 | -15% > gap widening |
| 4 | 15% or more gap closure (below attainment goal) | -4 | -11.25% > to \geq -15% |
| 3 | 11.25% \leq to <15% | -3 | -7.5% > to \geq -11.25% |
| 2 | 7.5% \leq to <11.25% | -2 | -3.75% > to \geq -7.5% |
| 1 | 3.75% \leq to <7.5% | -1 | 0 > to \geq -3.75% |
| 0 | 0% \leq to <3.75% | | |

Negative points are assigned for decreasing quality (“gap widening”). Gap widening of -15 percent or more is given -5 points, with -4 points assigned for $\frac{3}{4}$ - 1 of -15 percent gap widening, -3 points for $\frac{1}{2}$ - $\frac{3}{4}$, -2 points for $\frac{1}{4}$ - $\frac{1}{2}$, and -1 point for 0 - $\frac{1}{4}$ of -15 percent gap widening.

Points are weighted for individual measures and sub-measures as specified in Tables 3-7 above.

To address concerns about small year-to-year changes creating large absolute point assignments as plans near their attainment goal on a particular measure, ICHP has incorporated a 5 percent hold-harmless zone below the attainment goal. If a plan is within 5 percent of the attainment goal³ and incurs a year-to-year decline of 5 percent or less⁴ on the given measure, no negative points are assigned for that measure. For example, if the attainment goal for a measure is 82 percent, a plan that starts at or above 82 percent - 4.1 percentage points = 77.9 percent (note: the -4.1 percentage points equals -5 percent x 82 percent) will not receive negative points for a decline of 5 percent or less. So, if a plan begins at 80 percent (within 5 percent or 4.1 percentage points of the attainment goal), that plan can incur an incremental decrease in quality of up to -4 percentage points (80 percent x -5 percent = -4 percentage points) without earning negative points. In other words, the plan could decline to 76 percent without penalty.

Adjusted Points

The raw points for each plan are adjusted for both plan size and the plan’s number of missing measures. An adjustment for plan size is necessary because the amount of dollars at-risk varies directly with a plan’s capitation revenue while raw points are assigned solely based on incremental quality improvement. Consequently, without size adjustment, the P4Q impact will vary with plan size. Smaller plans will face greater risk (i.e., a greater range of positive and negative fiscal impacts) than larger plans. An adjustment for each plan’s number of missing measures is necessary because dollar allocations based on a plan’s total points will, all other factors constant, be smaller in absolute terms given missing measures. Missing measures can arise for a number of reasons, including failure to report data on a measure or because the number of eligibles for a measure falls below the minimum of 30.

The plan size adjustment factor equals the ratio of the plan’s program market share (plan revenues divided by total program revenues) to the inverse of the number of plans in the program. This ratio is then multiplied by each plan’s raw points score to produce the plan’s size-adjusted points score.

The intuition behind this plan size adjustment factor can most easily be understood by a numerical example. Keep in mind that the numerator (the plan market share) is the proportion of the total at-risk revenue pool accounted for by the plan while the denominator (the inverse of the number of plans in the program) is the expected proportion of the total points pool earned by the plan. Consequently, the plan size adjustment factor shows the relative size of (1) the plan’s expected contribution to the revenue

³ The phrase “within 5 percent of the attainment goal” refers to a plan starting at 95 percent or more of the attainment goal for a HEDIS measure.

⁴ The phrase “incurs a year-to-year decline of 5% or less” refers to the plan’s second year measure being 95% - 100% of the first year measure.

pool in relation to (2) the plan's expected contribution to the points pool. So, for example, suppose that Plan A has a 10 percent program market share and that there are 10 plans in the program. Plan A's plan size adjustment factor then equals $.10/(1/10)$ or 1.0, reflecting the fact that Plan A's contribution to the total revenue pool exactly equals its expected contribution to the points pool. Consequently, Plan A's raw points will exactly equal its size-adjusted points.

The situation is different for larger and smaller plans. Suppose that Plan B's market share is 20 percent. Plan B's plan size adjustment factor is then $.20/(1/10)$ or 2.0. Plan B's expected contribution to the revenue pool is twice its expected contribution to the points pool, so Plan B's raw points must be multiplied by two to yield adjusted points in line with the plan's contribution to the revenue pool. If Plan C's market share is 5 percent, Plan C's plan size adjustment factor is $.05/(1/10)$ or 0.5. Plan C's adjusted points will be one-half its raw points.

In summary, the plan size adjustment factor adjusts the plan's raw points total to ensure that its adjusted points total is commensurate with its relative contribution to the revenue pool.

The final adjustment to the points is for any missing measures. The missing measure adjustment factor is the number of weighted measures in the P4Q program divided by the weighted number of non-missing measures (i.e., the weighted measures available) for the individual plan. So, for example, if a program has four total weighted measures and Plan D is missing one weighted measure, Plan D's missing measure adjustment factor is $4/3$ or 1.33. Plan D's final adjusted total positive and negative points are calculated by multiplying its size-adjusted positive and negative points totals by the missing measure adjustment factor.

Converting Points to Dollars

This section presents the approach used to convert the quality points calculated above to dollar amounts for the Texas health plans. The conversion approach used for Texas dental plans is somewhat different, and is outlined in Appendix 2 to this report.

ICHP uses each plan's total positive and negative adjusted quality points as the basis for calculating that plan's P4Q positive and negative dollar amounts. Positive points earned through improvements in quality as shown by gap closure determine the dollar amounts paid to each plan while negative points assigned because of decrements in quality as evidenced by gap widening determine the dollar amounts that each plan must pay into the P4Q program. This is done in such a way as to ensure fiscal balance at the program level, i.e., the dollars paid by the plans into the P4Q program exactly equals the dollars paid out by the P4Q program to the plans. The total amount paid in and total amount paid out equal the specified percent of total program capitation revenues. Also, the net impact of the P4Q program on each plan is presently capped at +4 percent and -4 percent maximum gain and loss, respectively.

To ensure fiscal balance, ICHP calculates and applies a separate dollar amount per adjusted positive point and dollar amount per adjusted negative point for each program. ICHP then uses these dollars per positive (negative) point in conjunction with the plan's positive (negative) adjusted point totals to calculate the dollars paid to (paid by) the health plan. The steps in this process are: (1) sum individual plan capitation revenues across all plans to produce total program capitation revenues, (2) divide the specified percent of total program capitation revenues by adjusted positive (negative) points summed across all program measures and plans to obtain the dollars per positive (negative) point for the program, and (3) calculate the dollar amounts paid to (paid by) the plan as the product of dollars per positive (negative) point and total plan positive (negative) points. The net dollar impact on the plan is then the positive dollars paid to the plan based on the plans' adjusted positive point total minus the negative dollars paid by the plan based on the plan's adjusted negative point total.

If no plan earns more than +4 percent of its capitation revenues and no plan pays more than -4 percent of its capitation revenues, the positive and negative dollar allocations as calculated above stand as final amounts. However, if any plan earns more than +4 percent of its capitation revenues or pays more than -4 percent of its capitation revenues, the fiscal impact on such plans is capped at an absolute 4 percent of plan revenues. The net dollar amounts outside this absolute 4 percent cap (dollars above the +4 percent cap minus the dollars below the -4 percent cap) are summed across all plans to produce the net dollar amount for the program beyond the cap. This total dollar amount beyond the cap is then distributed to the plans within the absolute 4 percent cap in proportion to plan size so as to minimize the impact of the capping on the P4Q results. If this distribution causes any plan to exceed the absolute 4 percent cap, the capping process is repeated until all plans fall within the absolute 4 percent cap.

Issues in P4Q Design and Implementation

Risk Adjustment

HHSC, the EQRO, and the health plans have had numerous discussions about the appropriateness of risk adjustment and the best approach to conducting risk adjustment for the Texas P4Q program.

Conceptually, risk adjustment becomes attractive whenever enrollees distribute themselves across plans in a non-random way such that certain plans have enrollee pools with higher or lower claims or quality experience **due to factors beyond the plan's control**. Such factors thought to be beyond the plan's control commonly include enrollee health status or case-mix, demographics, and geographic location. There is no universal agreement about which factors that influence quality are beyond the plan's control, how such factors should best be measured, and how the effects of such factors on plan performance should be estimated. This lack of agreement exists in large part because (1) the factors and mechanisms that influence quality of care are imperfectly understood, (2) the ability of the plan to influence the factors and to alter the mechanisms is not always clear, and (3) many factors that influence quality are likely only partially modifiable by plans.

The goal in conducting such risk adjustment is not to develop a comprehensive statistical model of the determinants of quality or outcome. Rather, the purpose of risk adjustment is to remove the agreed-upon sources of differences in outcomes that are beyond the plan's control so that attention can be focused on the remaining differences that reflect plan performance.

Although the HEDIS measure used in P4Q are not presently risk-adjusted, we plan to continue discussions with HHSC and the MCOs concerning the choice of risk adjustors and the appropriateness of risk adjustment as the P4Q program evolves.

Statistical Significance in Pay-for-Quality (P4Q)

Classical statistical significance testing is used in scientific research when seeking to make an inference about a population from information contained in a sample. Such testing seeks to ensure that the observed results are real and not merely the result of random sampling variation. Statistical testing is set up to minimize the probability of a Type I error, where the null hypothesis is falsely rejected in favor of the alternative hypothesis when the null hypothesis is actually true. The probability of a Type II error (failing to reject the null hypothesis when it is indeed false) is left uncontrolled in this approach, and is therefore typically larger than the probability of a Type I error.

Such statistical significance testing is appropriate when the cost of committing a Type I error is substantially larger than a Type II error. In science, the cost of falsely rejecting the null hypothesis in favor of the alternative (a Type I error) is arguably higher than falsely failing to reject the null (a Type II error). Why? Because when the null is rejected, the alternative hypothesis becomes part of scientific knowledge. Because scientific knowledge should not be contaminated by false conclusions, researchers

concentrate on minimizing Type I errors. Better that a research study yield inconclusive results than endorse an incorrect result.

Incentivizing health plans through a P4Q program is in some ways similar to scientific research in that a sample is frequently being used to make an inference about a population. Consequently, an observed difference in a quality measure in a P4Q program may be real or it may be a result of random variation. It is tempting, therefore, to apply standard statistical significance testing to P4Q measures to identify those differences that fail to achieve statistical significance.

Before conducting such testing, however, it is useful to think about the relative costliness of Type I and Type II errors in P4Q programs. A Type I error in the P4Q context refers to falsely rewarding or penalizing a health plan when the criteria for rewards or penalties are not met. A Type II error in the P4Q context refers to failing to reward or penalize a plan when the incentive criteria are indeed met.

Which type of error is more costly in the context of P4Q? Is it more costly to falsely reward or penalize a plan, or is it more costly to withhold rewards or penalties when they have been earned? The answer is not as clear-cut as in scientific research, but it seems reasonable to suggest that the two types of errors are on more equal footing when incentivizing health plans than when pursuing scientific knowledge.

It could be argued that standard capitation payment without quality incentives implicitly places a higher relative weight on Type I-like errors compared to Type II-like errors in much the same way as scientific inquiry. When higher quality is valued, plans vary in quality, and plans are paid independently of their quality, payors commit a Type I-like error by failing to incentivize when incentives would be appropriate. Viewed in this light, P4Q programs are designed to increase the emphasis on Type II errors by seeking to recognize quality differences (at the cost of possibly recognizing non-existent quality differences) compared to payment schemes that do not recognize quality. In other words, P4Q programs place greater weight on failing to recognize quality differences when they do indeed exist.

This discussion should not be misconstrued as arguing that statistical testing is irrelevant to P4Q programs. Clearly, random variation is present in these P4Q measures and such variation needs to be recognized and evaluated. However, the uncritical application of classical statistical testing with its emphasis on Type I errors is not advisable. Similarly, Texas has appropriately rejected standard capitation payment without quality incentives and the implicit emphasis on Type I errors inherent in such a system. The appropriate role of statistical testing in P4Q programs involves a more balanced approach to Type I and Type II errors than what is achieved in either pure scientific research or standard capitation payment systems.

Intra-Year Updates

Plans have expressed a desire to have the ability to track their performance in the P4Q program during the course of the year in order to plan for the fiscal impact of the P4Q program and to accrue funds for any necessary recoupment. The EQRO will support health plans in their tracking by producing intra-year PPE expenditure reports at the mid-year point, and possibly quarterly if that proves feasible. For the HEDIS measures, the EQRO will produce interim reports in an effort to assist plans in their tracking efforts for these measures. The EQRO also plans to provide intra-year updates to the dental plans as well.

Time lags in encounter reporting will inevitably limit the degree to which these interim reports reflect real-time health plan experience with these measures.

Appendix 1. Gap Closure and the Partial Adjustment Model

The incremental improvement approach to P4Q outlined in this document is based on the concept of “gap closure.” This concept of gap closure is adapted from an econometric model of partial adjustment to equilibrium.⁵ A partial adjustment model consists of two equations, the first describing the determinants of the equilibrium value of the dependent variable, Y^* , and the second describing the process of adjustment over time to this equilibrium. Letting t index time and i index the individual cross-sectional unit (i.e., the health plan in this context), these two equations are:

$$(1) Y_{it}^* = \alpha + \beta X_{it} + \varepsilon_{it}$$

$$\text{where } \varepsilon \sim N(0, \sigma^2)$$

$$(2) Y_{it} - Y_{it-1} = \gamma(Y_{it}^* - Y_{it-1})$$

$$\text{where } 0 \leq \gamma \leq 1$$

Equation (1) is a simple linear regression showing that the equilibrium level of Y , Y^* , is based on a regressor X (e.g., case-mix) and a random disturbance ε that is normally distributed with mean zero and constant variance, σ^2 . Equation (2) shows the actual adjustment in Y between periods $t-1$ and t . This adjustment is assumed to be a constant fraction, γ , of the gap between the equilibrium and current levels of Y .

Inserting equation (1) into equation (2) and rearranging gives an estimable version of the partial adjustment model where the individual coefficients are readily identified:

$$Y_{it} = \alpha\gamma + \beta\gamma X_{it} + (1 - \gamma)Y_{it-1} + \gamma\varepsilon_{it}$$

In the P4Q approach, no attempt has been made to estimate a partial adjustment model from the data, but γ , the target gap closure, has been specified as 0.15. Also, the attainment goal corresponds to Y_{it}^* in the partial adjustment context.

This partial adjustment approach can be viewed as a generalization of a mean reversion or regression to the mean model. Rather than positing an equilibrium level of Y that can differ from the mean, the observed value of Y is assumed to vary randomly around its mean:

⁵ Dougherty, Christopher (2012) EC220 - Introduction to econometrics (chapter 11). [Teaching Resource: London School of Economics, <http://learningresources.lse.ac.uk/137/>]

$$Y_{it} = \bar{Y} + \varepsilon_{it}$$

where $\varepsilon \sim N(0, \sigma^2)$

$$Y_{it} - Y_{it-1} = \varepsilon_{it} - \varepsilon_{it-1}$$

There is no partial adjustment to equilibrium under mean reversion. Year-to-year changes are random and centered around zero. However, Y will tend to gravitate from values above or below the mean back to the mean. For example, an extreme value of Y in t-1 will likely be followed by a value closer to the mean in period t since extreme values of Y correspond to extreme values of ε . Since ε is random and draws are assumed independent, the value of ε in the next period is likely to be closer to zero and Y will be correspondingly closer to \bar{Y} .

In this context, the mean reversion model is a special case of the partial adjustment model where $Y_{it}^* = \bar{Y} + \varepsilon_{it}$ and $\gamma = 1$.

Appendix 2. Dental P4Q Earn Back Model

The P4Q approach for the two dental plans in Texas uses many of the same concepts as the P4Q approach for the health plans, including the same incremental improvement approach and points assignments based on gap closure. However, the approach to converting points to dollars is different. This appendix presents the points-to-dollars method used for the Texas dental P4Q program.

Because dental managed care in Texas presently consists of only two plans, the decision was made to give each plan an opportunity to “earn back” their 2 percent of revenue based on their quality points.

To be eligible to earn back any of their 2 percent of revenues, a dental plan must have more positive than negative points in absolute terms. For example, a plan that earns +9 positive points and -11 negative points would have a negative net point total and would not be eligible to earn back any of their 2 percent of revenues. Another plan that earns +15 points and -10 points would have a positive net point total and would be eligible to earn back some of their 2 percent of revenues.

Given positive net points, the amount earned back by a dental plan is based on the plan’s positive points as a percentage of the maximum possible positive points. A plan which earns 80 percent or more of the maximum possible positive points will recoup all of its 2 percent of revenue. Plans that earn less than 80 percent of the maximum possible positive points (and have positive net points) will earn back a percentage of their 2 percent P4Q revenue equal to the percentage of 80 percent of maximum positive points represented by their positive point total.

For example, a plan’s maximum point total with four P4Q dental measures is 4 measures x maximum 5 points /measure or +20 points. A plan with a net positive points total that earns 80 percent or more of this maximum (80 percent x 20 points = 16 points in this example) earns back their full 2 percent of revenues. Suppose that a plan has net positive points but only earns +8 positive points. These 8 positive points represent exactly one-half of 80 percent of the maximum positive points. Consequently, this plan earns back exactly one-half of its 2 percent of revenues, or 1 percent of revenues, and is left paying 1 percent of their revenues (-2 percent + 1 percent earn back = -1 percent net).